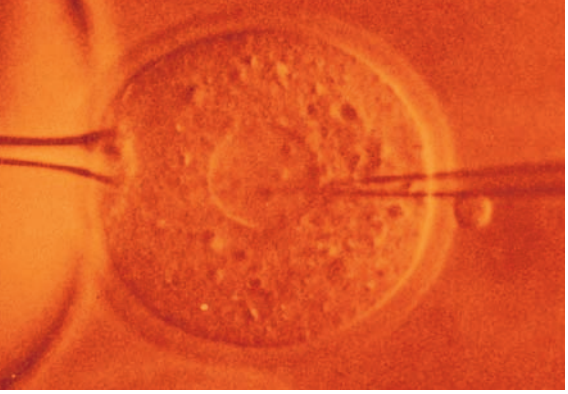


Nucleic Acids, Gene Expression, and Recombinant DNA Technology

CHAPTER 5

1 Nucleotides and Nucleic Acids
- A. Nucleotides, Nucleosides, and Bases
- B. The Chemical Structures of DNA and RNA

2 DNA Is the Carrier of Genetic Information
- A. Transforming Principle Is DNA
- B. The Hereditary Molecule of Many Bacteriophages Is DNA

3 Double Helical DNA
- A. The Watson–Crick Structure: B-DNA
- B. DNA Is Semiconservatively Replicated
- C. Denaturation and Renaturation
- D. The Size of DNA

4 Gene Expression and Replication: An Overview
- A. RNA Synthesis: Transcription
- B. Protein Synthesis: Translation
- C. DNA Replication

5 Molecular Cloning
- A. Restriction Endonucleases
- B. Cloning Vectors
- C. Gene Manipulation
- D. The Identification of Specific DNA Sequences: Southern Blotting
- E. Genomic Libraries
- F. The Polymerase Chain Reaction
- G. Production of Proteins
- H. Transgenic Organisms and Gene Therapy
- I. Social, Ethical, and Legal Considerations

Knowledge of how genes are expressed and how they can be manipulated is becoming increasingly important for understanding nearly every aspect of biochemistry. Consequently, although we do not undertake a detailed discussion of these processes until Part V of this textbook, we outline their general principles in this chapter. We do so by describing the chemical structures of nucleic acids, how we have come to know that DNA is the carrier of genetic information, the structure of the major form of DNA, and the general principles of how the information in genes directs the synthesis of RNA and proteins (how genes are expressed) and how DNA is replicated. The chapter ends with a discussion of how DNA is experimentally manipulated and expressed, processes that are collectively referred to as genetic engineering. These processes have revolutionized the practice of biochemistry.

1 NUCLEOTIDES AND NUCLEIC ACIDS

Nucleotides and their derivatives are biologically ubiquitous substances that participate in nearly all biochemical processes:

1. They form the monomeric units of nucleic acids and thereby play central roles in both the storage and the expression of genetic information.

2. Nucleoside triphosphates, most conspicuously ATP (Section 1-3B), are the "energy-rich" end products of the majority of energy-releasing pathways and the substances whose utilization drives most energy-requiring processes.

3. Most metabolic pathways are regulated, at least in part, by the levels of nucleotides such as ATP and ADP. Moreover, certain nucleotides, as we shall see, function as intracellular signals that regulate the activities of numerous metabolic processes.

4. Nucleotide derivatives, such as **nicotinamide adenine dinucleotide** (Section 13-2A), **flavin adenine dinucleotide** (Section 16-2C), and **coenzyme A** (Section 21-2), are required participants in many enzymatic reactions.

5. As components of the enzymelike nucleic acids known as **ribozymes,** nucleotides have important catalytic activities themselves.

A. *Nucleotides, Nucleosides, and Bases*

Nucleotides are phosphate esters of a five-carbon sugar (which is therefore known as a ***pentose;*** *Section 11-1A) in which a nitrogenous base is covalently linked to C1′ of the sugar residue.* In **ribonucleotides** (Fig. 5-1*a*), the monomeric units of RNA, the pentose is **D-ribose,** whereas in **deoxyribonucleotides** (or just **deoxynucleotides;** Fig. 5-1*b*),

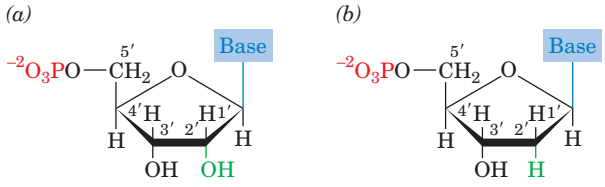
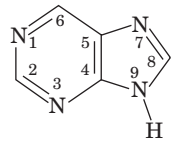


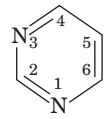
Figure 5-1 **Chemical structures of (*a*) ribonucleotides and (*b*) deoxyribonucleotides.**

the monomeric units of DNA, the pentose is **2′-deoxy-D-ribose** (note that the "primed" numbers refer to the atoms of the ribose residue; "unprimed" numbers refer to atoms of the nitrogenous base). The phosphate group may be bonded to C5′ of the pentose to form a **5′-nucleotide** (Fig. 5-1) or to its C3′ to form a **3′-nucleotide.** If the phosphate group is absent, the compound is known as a **nucleoside.** A 5′-nucleotide, for example, may therefore be referred to as a **nucleoside-5′-phosphate.** In all naturally occurring nucleotides and nucleosides, the bond linking the nitrogenous base to the pentose C1′ atom (which is called a glycosidic bond; Section 11-1Ca) extends from the same side of the ribose ring as does the C4′—C5′ bond (the so-called β configuration; Section 11-1Ba) rather than from the opposite side (the α configuration). Note that nucleotide phosphate groups are doubly ionized at physiological pH's; that is, *nucleotides are moderately strong acids*.

The nitrogenous bases are planar, aromatic, heterocyclic molecules which, for the most part, are derivatives of either ***purine*** *or* ***pyrimidine.***



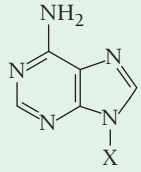
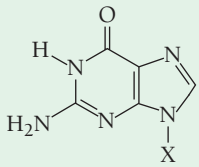
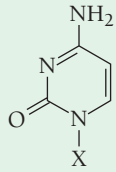
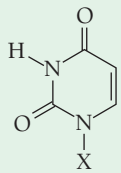
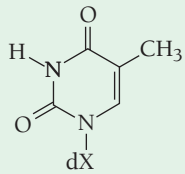
Purine



Pyrimidine

The structures, names, and abbreviations of the common bases, nucleosides, and nucleotides are given in Table 5-1. The major purine components of nucleic acids are **adenine** and **guanine** residues; the major pyrimidine residues are those of **cytosine, uracil** (which occurs mainly in RNA), and **thymine** (5-methyluracil, which occurs mainly in DNA). The purines form glycosidic bonds to ribose via

Table 5-1 **Names and Abbreviations of Nucleic Acid Bases, Nucleosides, and Nucleotides**

Base Formula	Base (X = H)	Nucleoside (X = ribose[a])	Nucleotide[b] (X = ribose phosphate[a])
	Adenine Ade A	Adenosine Ado A	Adenylic acid Adenosine monophosphate AMP
	Guanine Gua G	Guanosine Guo G	Guanylic acid Guanosine monophosphate GMP
	Cytosine Cyt C	Cytidine Cyd C	Cytidylic acid Cytidine monophosphate CMP
	Uracil Ura U	Uridine Urd U	Uridylic acid Uridine monophosphate UMP
	Thymine Thy T	Deoxythymidine dThd dT	Deoxythymidylic acid Deoxythymidine monophosphate dTMP

[a]The presence of a 2′-deoxyribose unit in place of ribose, as occurs in DNA, is implied by the prefixes "deoxy" or "d." For example, the deoxynucleoside of adenine is deoxyadenosine or dA. However, for thymine-containing residues, which rarely occur in RNA, the prefix is redundant and may be dropped. The presence of a ribose unit may be explicitly implied by the prefixes "ribo" or "r." Thus the ribonucleotide of thymine is ribothymidine or rT.
[b]The position of the phosphate group in a nucleotide may be explicitly specified as in, for example, 3′-AMP and 5′-GMP.

their N9 atoms, whereas pyrimidines do so through their N1 atoms (note that purines and pyrimidines have dissimilar atom numbering schemes).

B. *The Chemical Structures of DNA and RNA*

The chemical structures of the nucleic acids were elucidated by the early 1950s largely through the efforts of Phoebus Levene, followed by the work of Alexander Todd. *Nucleic acids are, with few exceptions, linear polymers of nucleotides whose phosphate groups bridge the 3′ and 5′ positions of successive sugar residues (e.g., Fig. 5-2).* The phosphates of these **polynucleotides,** the **phosphodiester** groups, are acidic, so that, *at physiological pH's, nucleic acids are polyanions.* Polynucleotides have directionality, that is, each has a **3′ end** (the end whose C3′ atom is not linked to a neighboring nucleotide) and a **5′ end** (the end whose C5′ atom is not linked to a neighboring nucleotide).

a. DNA's Base Composition Is Governed by Chargaff's Rules

DNA has equal numbers of adenine and thymine residues (A = T) and equal numbers of guanine and cytosine residues (G = C). These relationships, known as **Chargaff's rules,** were discovered in the late 1940s by Erwin Chargaff, who first devised reliable quantitative methods for the separation and analysis of DNA hydrolysates. Chargaff also found that the base composition of DNA from a given organism is characteristic of that organism; that is, it is independent of the tissue from which the DNA is taken as well as the organism's age, its nutritional state, or any other environmental factor. The structural basis for Chargaff's rules is that in double-stranded DNA, G is always hydrogen bonded (forms a **base pair**) with C, whereas A always forms a base pair with T (Fig. 1-16).

DNA's base composition varies widely among different organisms. It ranges from ~25% to 75% G + C in different

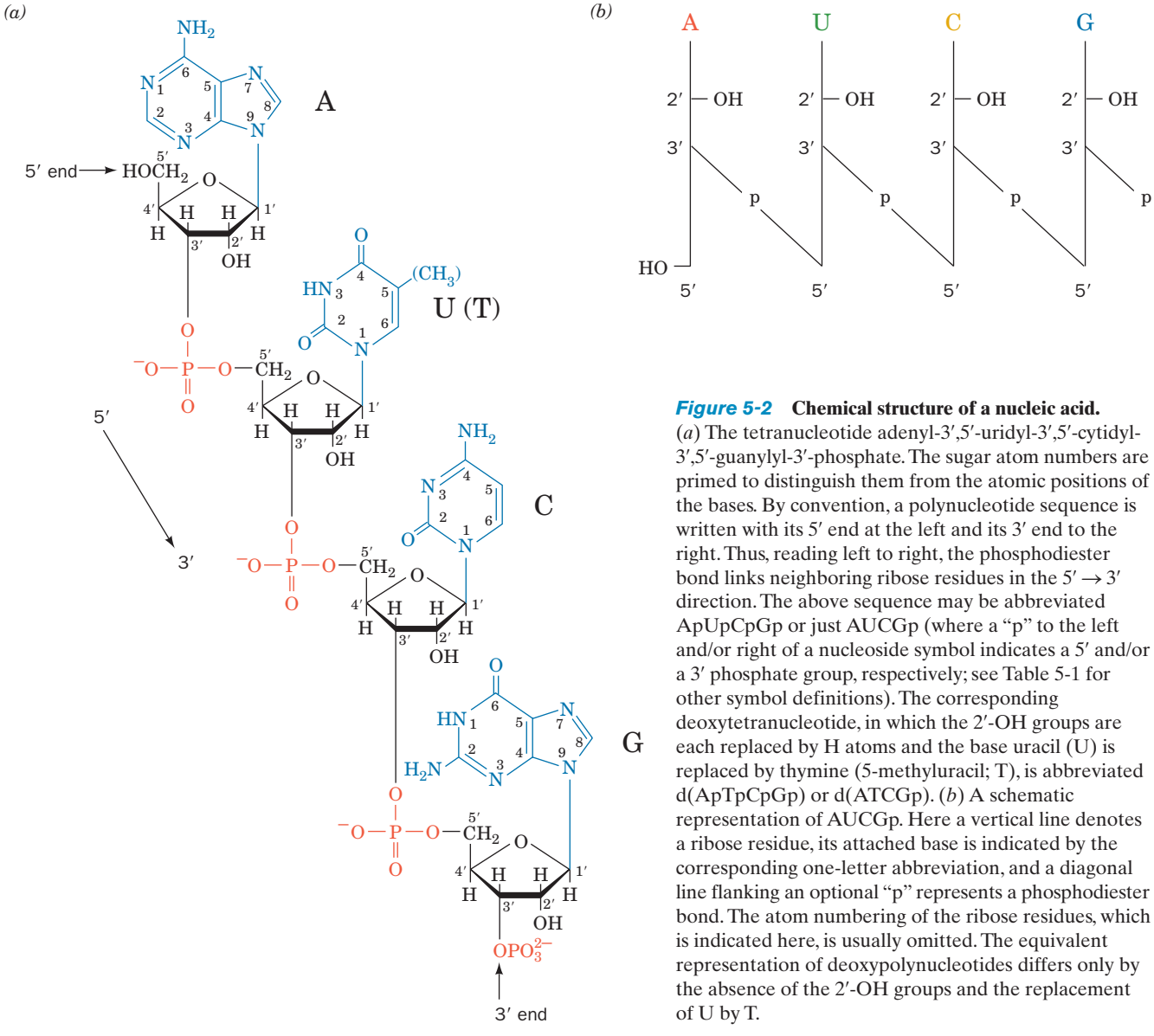


Figure 5-2 **Chemical structure of a nucleic acid.** (*a*) The tetranucleotide adenyl-3′,5′-uridyl-3′,5′-cytidyl-3′,5′-guanylyl-3′-phosphate. The sugar atom numbers are primed to distinguish them from the atomic positions of the bases. By convention, a polynucleotide sequence is written with its 5′ end at the left and its 3′ end to the right. Thus, reading left to right, the phosphodiester bond links neighboring ribose residues in the 5′ → 3′ direction. The above sequence may be abbreviated ApUpCpGp or just AUCGp (where a "p" to the left and/or right of a nucleoside symbol indicates a 5′ and/or a 3′ phosphate group, respectively; see Table 5-1 for other symbol definitions). The corresponding deoxytetranucleotide, in which the 2′-OH groups are each replaced by H atoms and the base uracil (U) is replaced by thymine (5-methyluracil; T), is abbreviated d(ApTpCpGp) or d(ATCGp). (*b*) A schematic representation of AUCGp. Here a vertical line denotes a ribose residue, its attached base is indicated by the corresponding one-letter abbreviation, and a diagonal line flanking an optional "p" represents a phosphodiester bond. The atom numbering of the ribose residues, which is indicated here, is usually omitted. The equivalent representation of deoxypolynucleotides differs only by the absence of the 2′-OH groups and the replacement of U by T.